Chapter XXXII

gLibrary/DRI: A Grid-Based Platform to Host Multiple Repositories for Digital Content

**ABSTRACT**

This chapter presents the gLibrary/DRI (Digital Repositories Infrastructure) platform. The main goal of the platform is to reduce the cost in terms of time and effort that a repository provider spends to get its repository deployed. This is achieved by providing a common infrastructure and a set of mechanisms (APIs and specifications) that repository providers use to define the data model, the access to the content (by navigation trees and filters) and the storage model. It also provides an algorithm launching mechanism. gLibrary/DRI offers a generic way to provide all this functionality; however, providers can add specific behaviours to the default functions for their repositories. The architecture is Grid based. Two use cases are also presented: A mammograms repository example that provides clinicians with a tool that eases diagnostics process and an algorithmic repository based on the Poincare Surface Section.
INTRODUCTION

This chapter describes the gLibrary/Digital Repositories Infrastructure (gLibrary/DRI) developed jointly by Istituto Nazionale di Fisica Nucleare (INFN) at Catania (Italy) and Centro Extremeño de Tecnologías Avanzadas, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CETA-CIEMAT) at Trujillo (Spain). Repositories are digital stores that manage data and metadata providing their access to users. In this sense, gLibrary (gLibrary, 2007) was conceived as a Grid implementation of a digital repository that takes advantage of the Grid features (VO authentication, file catalogues, metadata services, etc) offering an easy-to-use service and a powerful system to handle digital assets. The gLibrary/DRI infrastructure starts from this concept and extends the platform by offering a general multi-repository environment designed to support and manage several repositories. A gLibrary/DRI repository is made of large digital content (as image files, video, etc) and metadata associated with it (annotations, descriptions, etc.) The main goal of gLibrary/DRI is to reduce the cost in terms of time and effort that a repository provider spends in order to get its repository fully deployed.

In this sense, it is possible to distinguish between repository providers and repository users. A repository provider represents his community by defining the nature of the repository that his community will use. A repository user is a member of such community who interacts with the repository (to query, access, update and modify its content, as well as launch algorithms). gLibrary/DRI offers a platform that allows repository providers to (1) define the data structure (the data model) hold by their repository and (2) provide, if needed, specific viewers for their particular type of data. It also allows repository users to manage their data according to the specifications and tools defined by their repository providers along with launching algorithms that act on the repository data.

gLibrary/DRI defines four APIs so that repository providers can describe the characteristics of their repository. First, the data model API specifies how repository providers describe their data. Second, the storage API defines the rules for describing their data to be stored. For instance, a repository for digitalized manuscripts might define their data model to be composed of a set of images representing the digitalization of the pages of a manuscript, and then a series of metadata describing the historical context of the manuscript (year of creation, authors, historical description, etc). Another example, a repository of mammograms, could contain the images representing the mammograms, and metadata describing clinician diagnoses on them. Once this issue is achieved, the repository provider can determine that images will be stored in the Grid storage elements (profiting from replication of large data chunks, etc.) and then eventually decide that the metadata will be stored in a federated database. Third, the user interface API allows repository providers to describe; (1) what tools are needed to view their data (a specific viewer that shows pages of a manuscript on a book-like fashion, or mammograms belonging to a patient in the same physical application window for visual comparison) and, (2) what are the navigational structures which allow users to navigate through the repository, for instance, manuscripts will be browsed through two trees: one by year and another one by content subject. Finally, an algorithm launching API allows repository providers to define which processes will be available for their users to launch selected repository content. For instance, those processes could be able to select a set of manuscript pages and launch an image improvement algorithm on each page, followed by a process to extract the text content of the pages.

In summary, gLibrary/DRI is a multi-repository platform encompassing the most important Grid features such as VO authentication, data federation and replication, high computing power, etc. drastically reducing the effort required by a repository provider to deploy its digital content.
Related Content

Computer Aided Modeling and Finite Element Analysis of Human Elbow

Assistive Technology and Rehabilitation Engineering
[www.igi-global.com/chapter/assistive-technology-rehabilitation-engineering/63391?camid=4v1a](www.igi-global.com/chapter/assistive-technology-rehabilitation-engineering/63391?camid=4v1a)

Use of Telemedicine Systems and Devices for Patient Monitoring
[www.igi-global.com/chapter/use-telemedicine-systems-devices-patient/26253?camid=4v1a](www.igi-global.com/chapter/use-telemedicine-systems-devices-patient/26253?camid=4v1a)

Convergence of Nanotechnology and Microbiology
[www.igi-global.com/chapter/convergence-of-nanotechnology-and-microbiology/186670?camid=4v1a](www.igi-global.com/chapter/convergence-of-nanotechnology-and-microbiology/186670?camid=4v1a)